

**UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF ILLINOIS**

KLEEN PRODUCTS LLC, individually and
on behalf of all others similarly situated,

Plaintiffs,

vs.

PACKAGING CORPORATION OF
AMERICA, et al.,

Defendants.

Case No. 1:10-cv-05711

Judge Milton I. Shadur

Magistrate Judge Nan R. Nolan

**PLAINTIFFS' REPLY MEMORANDUM OF LAW
FOR EVIDENTIARY HEARING**

I. INTRODUCTION

With the assistance of expert testimony, Plaintiffs will establish at the evidentiary hearing that there are critical deficiencies in Defendants' Boolean search string methodology and the process used in its development. Plaintiffs' experts will also detail how a content-based search methodology will efficiently maximize production of responsive documents in this case. They will further testify that Defendants' document retention and collection practices – including the fact that certain Defendants apparently permitted employees to *self-collect* their own responsive documents or otherwise failed to capture large sources of responsive information – weigh strongly in favor of a robust collection from all potential data storage locations based on corporate function, department, or work unit. In short, Defendants' collection process to date is inherently flawed and was imposed unilaterally; Plaintiffs have never seen many of the proposed search strings before they were filed as exhibits to Defendants' opening brief. In addition, because Defendants' document productions have not been organized in any meaningful way, with large portions devoid of metadata, the productions should be appropriately indexed. Finally, the Court should determine the time period for production based on the allegations in Plaintiffs' complaint, Plaintiffs' document requests, and relevant case law.

II. ARGUMENT

A. The Complexity and Stakes of This Case Requires a Content Search Of Data From All Sources Likely to Contain Responsive Information.

Plaintiffs' experts Dr. Lewis, Mr. Hanners, and Dr. Carol Tenny (a linguist) will testify regarding numerous deficiencies with Defendants' Boolean process, including the following:

- Defendants have not included obvious abbreviations for containerboard products used by their employees, for example "CTBD."
- Defendants' reporting of statistical results is logically incoherent and is reported in a selective fashion that falsely conveys a sense of statistical rigor. Consequently, it is not possible to draw any valid conclusions about the effectiveness of their procedure.

- Defendants excluded entire domains from the email they searched, but have only disclosed three such domains. One of the three disclosed domains, expedia.com, clearly may contain responsive information regarding travel itineraries and opportunity contacts between Defendants.
- Defendants' search strings do not capture numerous variations of key concepts. For example, Search String 8 (Defs.' 2/6/12 Br., Ex. 5 at 3 (ECF No. 288)) appears intended to obtain price-change related information, as requested in Plaintiffs' RPD Nos. 29-30, 32, and 34. *See* Defs.' 2/6/12 Br., Ex. 1. However, a simple internet search of Paper Age or Pulp and Paper's websites (two containerboard industry publications), indicates that the string would not capture responsive documents describing "pricing up," "prices rose," or prices "became more expensive."
- Individuals assessing the validity of Georgia Pacific's ("GP") search process were not "blind" to whether the results they reviewed were actually obtained by Defendants' search process – instead, reviewers labeled some documents "marginally responsive" and used other techniques that biased their results to indicate better performance.
- Defendants sampling frame is fatally flawed.
- Defects in Defendants' preservation and collection issues compound the many faults in their search validation process.

Dr. Lewis will testify that a central problem with GP's validation process is their choice of "sampling frame" – *i.e.*, the sub-set of documents GP used to validate the responsiveness rate of their Boolean search strings. Contrary to best practices, GP used a limited five-custodian data set (including one employee whose data was admittedly irrelevant),¹ restricted by a narrow time limitation, to validate its Boolean search strings, rather than the complete universe of data from all custodian and non-custodian sources. *See Reference Manual on Scientific Evid.* (2d ed.), at 241 (Federal Judicial Ctr. 2000), available at [http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/\\$file/sciman00.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf) ("A survey that provides information about a wholly irrelevant universe of respondents is itself irrelevant."). They then seek to extrapolate the result of their flawed test to the tainted universe of documents, without a valid basis to do so. Defendants'

¹ Though other Defendants claim to have "validated the search terms on their own respective ESI," they did so on a limited set of non-responsive information from an undisclosed number of custodians without including all applicable ESI from the appropriate temporal scope or data storage locations. *See* Defs.' 2/6/12 Br. at 10.

failure to collect data from the relevant time period, as well as their deficient preservation and collection efforts,² means that Defendants' validation process is fatally flawed.

Regardless of what "routinely used" search methods might be appropriate in a "common" case, *see* Defs.' 2/6/12 Br. at 17-18, Defendants cannot seriously dispute that for purposes of *this case*, Plaintiffs' proposed content-based search is superior, both in terms of the total percent of responsive documents obtained and capturing documents that may use "coded" language. *See* Pls.' 2/6/12 Br. at 6 (ECF No. 290) (CBAA typically obtains no less than 70 percent (worst case) while Boolean searching may obtain as few as 24 percent (best case); CBAA also reduces e-discovery costs). Defendants completely ignore the relative merits of CBAA versus Boolean searching and fail to present any comparison of the two methods.

Defendants claim that Boolean searching is the *sine qua non* of e-discovery and content-based searching is a "new, largely untested" search methodology. *See* Defs.' 2/6/12 Br. at 1. In fact, Boolean searching, based on an algebraic theory developed in the 1840s, is notorious for its deficiencies, whereas content-based searching, which has been around since the 1970s, is recognized as a superior e-discovery search tool by, among others, *Defendants' attorneys and e-discovery consultants*. *See* Pls.' 2/6/12 Br. at 7-10. Defendants' expert, Mr. Brown, will apparently testify that "he is not aware of other available document review methods that would achieve greater success in identifying potentially responsive ESI [than keyword searching]"

² Even where Defendants have preserved certain information, it was not included in the set of data upon which the search terms were "validated." For instance, Temple-Inland refuses to search more than the 60 days of email not automatically deleted from a custodian's email account, but they have 8 years (2003-2010) of email (and other documents) on backup tapes that could be restored and searched. Defs.' 2/6/12 Br., Ex. 21 at 2-3. Without this large volume of information included in the data set used to validate search terms, Temple-Inland simply cannot make any statistically valid statements about the effectiveness of its Boolean searches.

(Defs.’ 2/6/12 Br. at 18), which directly contradicts reports and press releases of KPMG, which employs Defendants’ other search methodology expert, Mr. Koch.³ *See* Pls.’ 12/13/11 Br. at 6.

Defendants also misrepresent Plaintiffs’ proposal as a “challenge [to] the use of [keyword] search terms *per se*.⁴” Defs.’ 2/6/12 Br. at 1. In fact, Plaintiffs have repeatedly emphasized that the circumstances of *this* complex antitrust class action should drive selection of an appropriate search methodology. *See, e.g.*, Pls.’ 2/6/12 Br. at 9-10. Accordingly, Defendants’ reliance on “default” e-discovery orders is entirely misplaced.⁴ *See* Defs.’ 2/6/12 Br. at 11.

Defendants miss the central focus of the evidentiary hearing: whether use of a Boolean search in this case will be as effective in locating responsive documents as compared to a content-based search. Their experts appear to ignore the evolution of information retrieval technology (while actively promote CBAA searching to their other clients) and will testify that Defendants’ Boolean keyword search process is “effective at identifying potentially responsive ESI.” Defs.’ 2/6/12 Br. at 17. This begs the obvious question: “effective” as compared to what?⁵

Defendants’ experts should not be permitted to argue that the self-created “prejudice” Defendants claim as a result of pursuing their Boolean search in the face of Plaintiffs’ early and

³ While Mr. Koch acknowledges that “KPMG and other companies have developed new analytical tools as alternatives to the use of search terms” (Defs.’ 2/6/12 Br. at 17), he claims that KPMG did not have such tools available until September 2011. This is simply not accurate. KPMG acquired content-based search capabilities in 2006 through its’ UK subsidiary. *See* Ex. A hereto (June 5, 2006 KPMG press release).

⁴ Notably, the drafting committee for one of those orders – the Federal Circuit eDiscovery model order – observed that “keyword searching against large volumes of data to find relevant information is a challenging, costly, and imperfect process.” *See* Fed. Cir. eDiscovery Model Order, *Discussion* at 3 (citation omitted), *available at* http://www.cafc.uscourts.gov/images/stories/announcements/Ediscovery_Model_Order.pdf.

⁵ Such an approach is akin to an argument that choosing a horse as a mode of transportation is acceptable because it is the best available horse, even though technology has evolved and a superior form of transportation – the automobile – is now available. In this case, by ignoring the advancement of information retrieval tools, Defendants’ experts are left arguing that they have tried to enhance or account for the numerous deficiencies of a Boolean search, but they cannot avoid the end result: they are not using the best of the alternatives available.

oft-repeated objections excuses them from making a proper search. *See* Defs.’ 2/6/12 Br. at 17 (“It would not make sense for Georgia-Pacific now to create and implement an entirely new process”); *id.* at 18 (“[I]t would be inappropriate for Georgia-Pacific now to create and implement an entirely new process”). Expert testimony is meant to inform the Court about technical subjects, not to make legal or factual arguments. Moreover, as these experts should be aware, Defendants’ investment in developing a Boolean keyword search would not be wasted simply because a content-based search is ultimately used. *See* Pls.’ 2/6/12 Br. at 12 & n.13.

Finally, contrary to Defendants’ suggestion that they engaged in good-faith and meaningful cooperation with Plaintiffs (*see* Defs.’ 2/6/12 Br. at 4, 12-13), they ignored both repeated requests to use content-based analytics and Plaintiffs’ pointed criticisms of their proposed search strings, instead opting to plow ahead unilaterally.⁶ Indeed, the Defendant-specific search strings attached to Defendants’ opening brief (Defs’ 2/6/12 Br., Exs. 6-11) were never disclosed previously to Plaintiffs.⁷

In the absence of cooperation regarding basic preservation, collection, and search obligations, the court may impose whatever remedy is best suited to resolve the issue. *Pippins v. KPMG LLP*, -- F.R.D. --, 2012 WL 370321, at *12 (S.D.N.Y. Feb. 3, 2012) (holding that KPMG had failed to meaningfully cooperate regarding the scope of its preservation obligations and requiring KPMG to preserve all information unless it cooperated in sampling potentially responsive hard drives). In this case, Plaintiffs first requested subject and department searches in December 2010 and within days of Defendants’ announcement that they would present Boolean

⁶ After Defendants’ August 2011 disclosure that they had begun developing Boolean search strings, Plaintiffs engaged in communications with Defendants to point out the many deficiencies of their draft search strings – but repeatedly noted that such comments were not a “complete, or final, submission on search terms or any search protocol.” Defs.’ 2/6/12 Br., Ex. 4 at 1. *See also* Ex. B hereto (Nov. 10, 2011 letter from Douglas A. Millen highlighting specific deficiencies in Defendants’ proposed search strings).

⁷ Only Exhibit 11, Weyerhaeuser’s search strings, is dated. Exhibit 11 is “as of 12/7/2011.”

search strings, Plaintiffs reiterated that they wanted Defendants to “use advanced analytics, instead of Boolean search strings . . . [because] better analytical tools are available.” Ex. C hereto (Aug. 8, 2011 email) (emphasis in original). Nevertheless, Plaintiffs understood their obligation to cooperate, consistent with Sedona Conference and Seventh Circuit standards, and in good faith sought to learn more about Defendants’ proposed search method.⁸ Unfortunately, Defendants frustrated these efforts by not providing Plaintiffs with basic information necessary to assess their search process, such as the hit counts for specific search strings or readily available reports from concept clustering apparently used by Defendants to enhance their search terms. To make matters worse, Defendants refused to ask for details or discuss in any way Plaintiffs’ proposed content-based search methodology until November 8, 2011, on the eve of the status conference scheduled with Judge Shadur and weeks after declaring that their Boolean search process was complete. *See* Ex. D hereto. Permitting Defendants to rely upon their ill-advised and speculative investment in a hotly disputed search methodology, while only engaging Plaintiffs *after the fact*, would reward them for paying lip service to the level of cooperation envisioned by the Sedona Conference and the Seventh Circuit Electronic Discovery Pilot Program.

B. A Search of All Relevant Data Storage Locations Based on Corporate Function, Department, or Work Unit is Essential Given the Clear Deficiencies in Defendants’ Document Collection Practices.

Inexplicably, some Defendants appear to have permitted key personnel to *self-collect* documents to be preserved from their own hard drives – a clear breach of standard e-discovery

⁸ For instance, Plaintiffs asked Defendants on May 31, 2011 about the CBAA search tools to be used in searching for responsive documents – prior to GP’s behind-the-scenes development of a Boolean search protocol. Defendants’ citation of Plaintiffs’ May 31, 2011, email as evidence that Plaintiffs have not “consistently rejected the use of search terms” (Defs.’ 2/6/12 Br. at 12 n.15) is tenuous given that this email so clearly references concept-based searching and reflects Plaintiffs’ efforts to engage in a cooperative search process. *See* Defs.’ 2/6/12 Br., Ex. 13.

practice,⁹ which is particularly egregious in a case alleging an illegal price-fixing conspiracy. For example, International Paper (“IP”) appears to have permitted its identified custodians to select which ESI would be collected for review from their hard drives, thus allowing these individuals – who may have participated in a price fixing conspiracy – to self-select what would and would not be collected. *See* Defs.’ 2/6/12 Br., Ex. 18. And IP is not alone in this practice. Mr. Hanners will testify that this deficiency warrants full and complete collection of all remaining information on a corporate function, department, or work unit basis. Further, Defendants’ preservation and collection deficiencies seriously corrupt the potential universe of documents subject to any search methodology – Boolean or CBAA.¹⁰ “A court - and more importantly, a litigant - is not required to simply accept whatever information management practices a party may have. A practice may be unreasonable, given responsibilities to third parties. While a party may design its information management practices to suit its business purposes, one of those business purposes must be accountability to third parties.” *Phillip Adams & Assocs., L.L.C. v Dell, Inc.*, 621 F. Supp. 2d 1173, 1193 (D. Utah 2009).

Defendants also inexplicably excluded certain key personnel as document custodians. For example, GP has refused Plaintiffs’ repeated requests to add President and Chief Operating Officer James Hannon as a document custodian. Even without the benefit of GP’s internal documents, Plaintiffs are aware that Mr. Hannon’s responsibilities at GP included containerboard

⁹ *Danis v. USN Commc’n, Inc.*, No. 98 C 7482, 2000 WL 1694325, at *1-*2, *38 (N.D. Ill. Oct. 23, 2000) (failure to audit and objectively assess custodian’s document preservation constitutes “extraordinarily poor judgment”); *accord Jones v. Bremen High Sch. Dist.* 228, No. 08 C 3548, 2010 WL 2106640, at *7 (N.D. Ill. May 25, 2010) (party was grossly negligent for relying on employees to preserve and identify relevant documents because “employees are often reluctant to reveal their mistakes or misdeeds”).

¹⁰ While Defendants should have every motivation to establish the sufficiency of their efforts to preserve relevant information, instead they provided flawed and incomplete letters (Defs.’ 2/6/12 Br., Exs. 16-22) regarding their preservation and collection practices and have uniformly either delayed or refused to provide additional information. Plaintiffs will obtain further information regarding deficiencies in Defendants’ preservation practices and file any appropriate spoliation motions at a later date.

pricing and profit. *See* Ex. E hereto (June 2007 *Pulp & Paper* article). Intentional exclusion of such an obvious custodian demonstrates how Defendants' custodian-based approach excludes important and relevant evidence. *See, e.g., In re Citric Acid Litig.*, 996 F. Supp. 951, 953 (N.D. Cal. 1998) (detailing price-fixing conspiracy effectuated by top executives known as "masters" and implemented by lower-level corporate officials known as "Sherpas"). A collection based on corporate function, department, or work unit eliminates reliance on custodian selection and is likely to result in substantially greater recall of responsive documents.

C. Glaring Deficiencies in Defendants' Document Collection Practices Negate Any Value From Their Metadata and Further Supports The Need For Defendants to Label Their Document Production.

Rule 34(b)(2)(E)(i)'s requirements are clear: Defendants must produce their documents as kept in the usual course of business or index their production to correspond to Plaintiffs' RPDs. Plaintiffs simply seek to hold Defendants to the rule. If records have been reorganized or do not maintain the original structure, they are not within Rule 34(b)(2)(E)(i)'s definition of "usual course of business." *See In re Sulfuric Acid Antitrust Litig.*, 231 F.R.D. 351, 363 (N.D. Ill. 2005); *Am. Int'l Specialty Lines Ins. Co. v. NWI-I, Inc.*, 240 F.R.D. 401, 410-11 (N.D. Ill. 2007). Defendants unquestionably have not produced documents as kept in the usual course of business as they cannot claim that they store their documents in the usual course of business as TIFF images in no apparent order, with only a portion of limited metadata fields.

Defendants have failed to produce metadata for most of the documents they have produced to date, thus negating any argument that metadata alone is sufficient to satisfy the "usual course of business" requirement. Mr. Hanners will provide testimony regarding failures by a number of Defendants to collect documents in a forensically sound manner that has resulted in corrupted or altered metadata fields. For instance, Mr. Hanners will testify that numerous

documents produced by Defendants have metadata that reflect modification dates in 2011 – months after Defendants’ litigation holds were supposedly put in place. Consequently, Defendants’ reliance on metadata to satisfy Rule 34(b)(2)(E)(i)’s usual course of business requirement is misplaced.

Because Defendants have not produced their documents as kept in the ordinary course of business, Defendants turn their attack on Rule 34(b)(2)(E)(i) itself. Defendants contention that Rule 34(b)(2)(E)(i) would require them to “code” documents for Plaintiffs is provocative, but false. *See* Defs.’ 2/6/12 Br. at 2, 16. Clearly, Plaintiffs are not asking Defendants to “code” documents based on level of importance. Plaintiffs merely ask that Defendants comply with Rule 34(b)(2)(E)(i) by providing indexes or by correlating documents to specific P1-RPD. It is unfair for Defendants to criticize Plaintiffs for providing a list of 29 subject matter categories to simplify Defendants’ organization of their documents; it was Defendants who *twice requested* such a list. *See* Pls.’ 2/6/12 Br. at 19. Further, while Defendants’ claim that Plaintiffs’ demand is “unprecedented” (Defs.’ 2/6/12 Br. at 7) some or all Defendants are presently providing indexes of their documents to the Department of Justice and Plaintiffs have offered to accept such indices in lieu of Rule 34’s labeling requirement. Pls.’ 2/6/12 Br. at 19.¹¹

D. The Court Should Resolve the Time Period and Scope of Search Issues.

Defendants’ argument that it is premature to consider the relevant discovery time period and scope of locations to be searched is illogical. *See* Defs.’ 2/6/12 Br. at 2. These are not ESI issues. For both issues, the Court need only look to Plaintiffs’ complaint, Federal Rule of Civil

¹¹ Defendants disingenuously argue that “the indexing ship has already sailed” because some documents have already been reviewed. Defendants decided to proceed without resolving Plaintiffs’ concerns regarding the proper search methodology at their own peril. Whether they have already reviewed “millions of pages” of documents is therefore irrelevant as they did so with full knowledge of Plaintiffs’ objections. *See* Defs.’ 2/6/12 Br. at 7.

Procedure 26, and relevant case law. *See* Pls.’ 12/13/11 Br. at 16-18; Pls.’ 2/6/12 Br. at 14-17. Far from being premature, a ruling on what information is relevant regarding these two issues will assist the parties in resolving any future disputes that may arise. In the absence of such a ruling, Plaintiffs are stuck where they have been for the past eight months – faced with claims by Defendants that a broader temporal scope or full search of data storage locations is irrelevant and the vague claim that “some” Defendants “may” face an undue burden if required to collect and produce the information Plaintiffs request. *See* Defs.’ 2/6/12 Br., Ex. 5-6.

III. CONCLUSION

When compared head-to-head, Plaintiffs’ content-based search methodology is far superior in this case to Defendants’ Boolean search methodology. Plaintiffs’ proposal is also in line with the broad discovery afforded in antitrust actions and with the principles of both the Sedona Conference and Seventh Circuit Electronic Discovery Pilot Program. Further, given the deficiencies in Defendants’ collection procedures, such as improperly excluding custodians and allowing self-selected custodians to *self-collect* their own documents, a robust corporate function, department, or work unit based collection procedure is necessary. Defendants must also comply with the clear requirements of Rule 34(b)(2)(E)(i) and index their production. Finally, the relevance of Plaintiffs’ requested time period and scope of data storage locations to be searched is clear and should be decided by the Court at this time.

Dated: February 13, 2012

Respectfully submitted,

/s/ Michael J. Freed

Michael J. Freed
Steven A. Kanner
Robert J. Wozniak
FREED KANNER LONDON
& MILLEN LLC
2201 Waukegan Road, Suite 130
Bannockburn, IL 60015
Tel: (224) 632-4500
Fax: (224) 632-4521

Daniel J. Megin
THE MOGIN LAW FIRM, P.C.
707 Broadway, Suite 1000
San Diego, CA 92101
Tel: (619) 687-6611
Fax: (619) 687-6610

Interim Co-Lead Class Counsel

W. Joseph Bruckner
Brian D. Clark
LOCKRIDGE GRINDAL NAUEN P.L.L.P.
100 Washington Avenue South
Suite 2200
Minneapolis, MN 55401
Tel: (612) 339-6900
Fax: (612) 339-0981

Walter W. Noss
SCOTT+SCOTT LLP
707 Broadway, Suite 1000
San Diego, CA 92101
Tel. (619) 233-4565
Fax: (619) 233-0508

Class Counsel